

## **Dados para a Transformação: Inteligência Artificial e Análise de Dados no Planejamento Socioambiental e redução do impacto de queimadas em 200 comunidades do estado do Pará**

### **1. Contextualização do Problema**

Problema - Alto índice de queimadas em municipalidades do estado do Pará

A análise de 200 comunidades do estado do Pará revelou uma possível correlação entre carência de infraestrutura e maior pressão ambiental. Observa-se que comunidades com menor presença de unidades de saúde e escolas tendem a registrar maiores índices de desmatamento e menor cobertura vegetal. Essa relação sugere que a ausência de serviços públicos básicos pode contribuir para maior vulnerabilidade ambiental, seja por falta de alternativas econômicas sustentáveis ou pela menor presença do Estado.

Há comunidades com desmatamento acima de 80%, possivelmente por desmatamento recente ou florestas isoladas. O acesso à água potável é desigual, com algumas em situação crítica (menos de 35%). A renda mensal varia de R\$412 a R\$2.992, refletindo forte desigualdade. A ocorrência de queimadas chega a 11 por ano em certos locais, gerando riscos ambientais e à saúde. A distância de até 100 km dos centros urbanos dificulta o acesso a serviços básicos.

### **2. Hipótese do Grupo**

Existe a hipótese de que, quanto maior o índice de desmatamento, menor é o acesso à água potável. Além disso, o número de queimadas por ano também pode contribuir para essa relação, como demonstram os gráficos de dispersão apresentados nas Figuras 1 e 2. Além dessas duas variáveis, outros parâmetros também podem influenciar negativamente o acesso à água potável.

GitHub com mais informações: <https://github.com/feliperafaelbarbosa/exploratory-analises-group20-a2/blob/main/data-exploratory-analysis.ipynb>

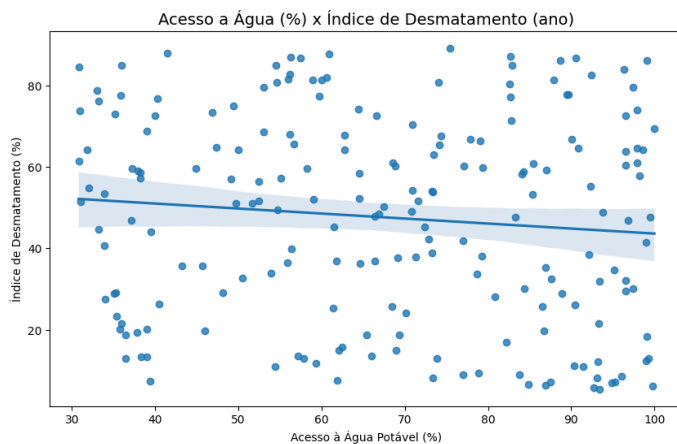


Figura 1: Acesso a água x Desmatamento

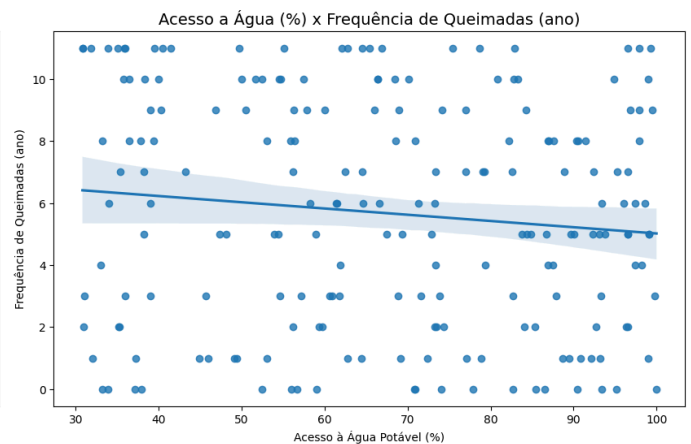


Figura 2: Acesso a água x Queimadas

### 3. Recomendação de Uso da IA

Uma abordagem recomendada é a criação de um mapa interativo utilizando a biblioteca Folium. Este mapa indicaria o nível de risco de incêndio variando entre os status: Baixo, Médio ou Alto Risco. A partir dessa visualização, seria possível formular hipóteses e previsões sobre a ocorrência de focos de incêndio nesses locais.

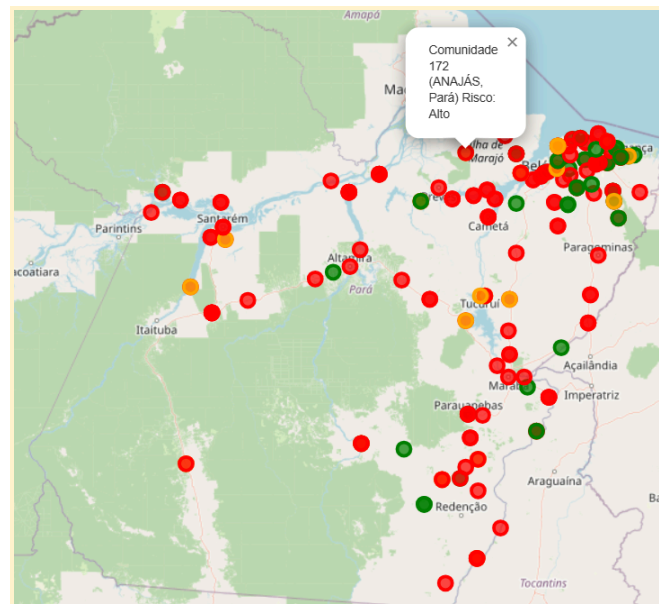


Figura 3: Distribuição em cores de risco de focos de incêndio por município no Pará.

Nota-se, a partir da análise do mapa, que a região do Nordeste paraense apresenta maiores focos de incêndios relatados no conjunto de dados do que o restante dos municípios do estado.

## 4. Representação dos Dados do Dataset

Algumas considerações a respeito da fidedignidade dos dados:

O dataset não apresenta dados faltantes nos campos numéricos e categóricos, como também não há a presença de outliers apresentado na Figura 4, o que contribuiu significativamente para a exploração dos dados.

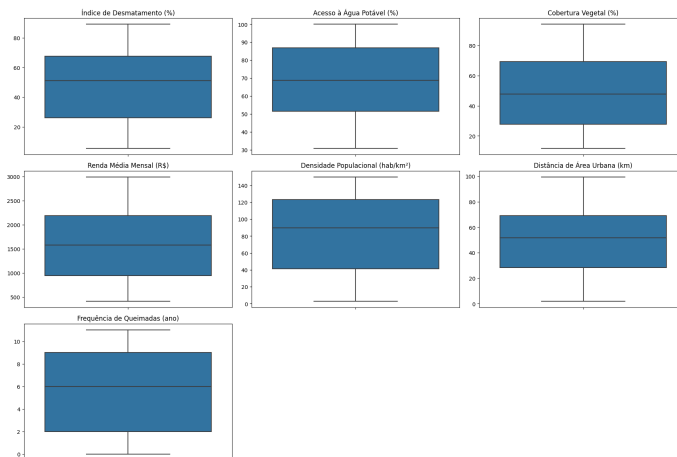


Figura 4: Explorando outliers em dados numéricos (ints e floats).

Os box plots (Figura 5 e 6) mostram a distribuição da frequência de queimadas anuais em regiões com e sem escolas (esquerda), e com e sem unidades de saúde (direita), indicando que a presença ou ausência dessas infraestruturas não parece ter um impacto significativo na mediana da frequência de queimadas.

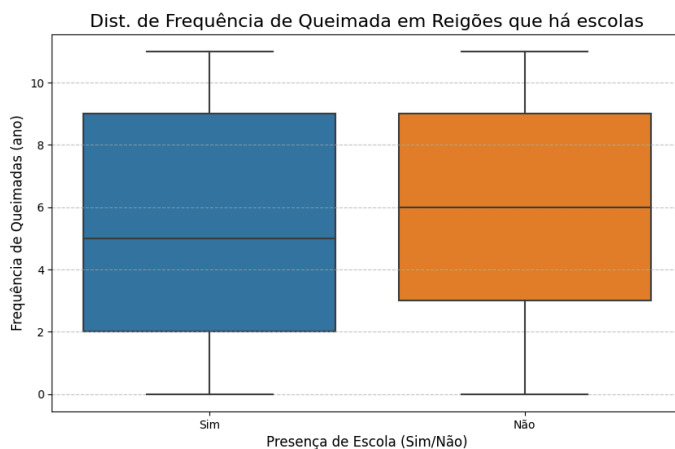


Figura 5: Queimadas x Presença de escolas

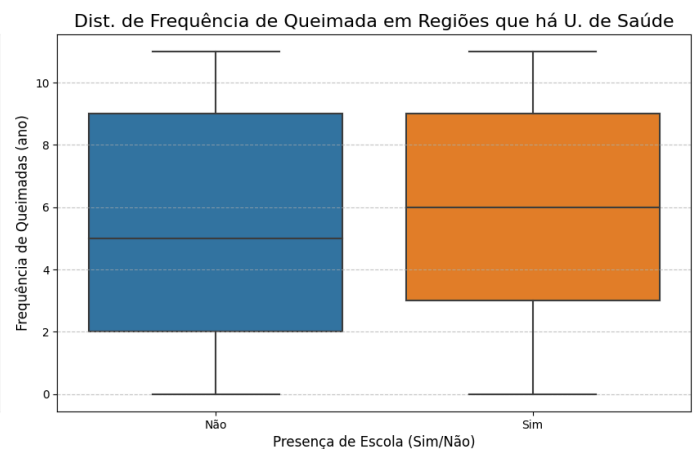


Figura 6: Queimadas x Presença de unidades de saúde

Além disso, a baixa correlação entre as variáveis apresenta um desafio significativo para a utilização dos dados na aferição de causalidade e na construção de modelos preditivos.

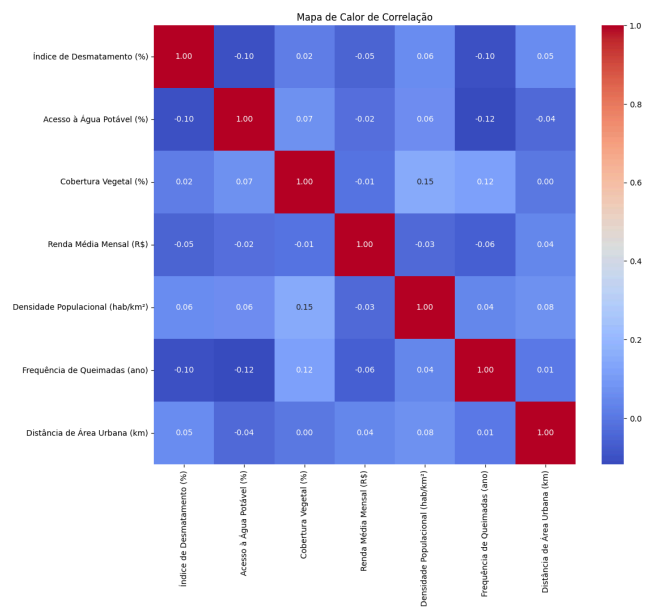


Figura 7: Mapa de calor de correlação entre as variáveis numéricas, os valores de correlação variaram entre -0.12 e 0.15.

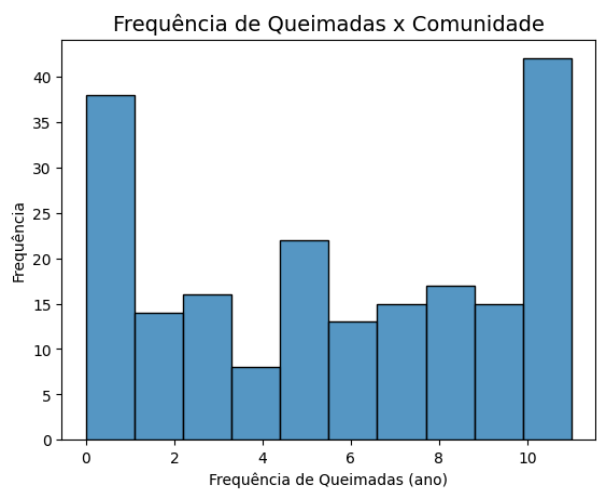


Figura 8: Distribuição da frequência de queimadas ao ano por comunidades.

Na Figura 8, o histograma da frequência anual de queimadas mostra uma distribuição bimodal, com picos notáveis em aproximadamente 0-1 e 10-11 queimadas por ano. Isso sugere que a maioria das regiões estudadas ou não apresenta queimadas ou sofre com uma alta frequência, com menos ocorrências nas faixas intermediárias.

Em suma, o dataset carece de metadados, pois não sabemos a fonte original, data de coleta, instituição responsável, comprometendo a veracidade dos fatos, o que pode ocasionar o seu mal uso para abordagens reais, como por exemplo: criação de políticas públicas e tomadas de decisões para implementação de projetos eficazes que combatem os efeitos causados pela degradação ambiental.

## **5. Identificação do Grupo**

### **Grupo 20**

Nome completo dos integrantes:

- 1. Felipe Barbosa